NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Purple and Blue text represents the responses from measure developers

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0058

Corresponding Measures:

De.2. Measure Title: Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (AAB)

Co.1.1. Measure Steward: National Committee for Quality Assurance

De.3. Brief Description of Measure: The percentage of episodes for members ages 3 months and older with a diagnosis of acute bronchitis/bronchiolitis that did not result in an antibiotic dispensing event.

1b.1. Developer Rationale: The vast majority of acute bronchitis cases are viral. Bacteria are detected in 1% to 10% of cases, and can include Bordetella pertussis, Chlamydia pneumoniae, and Mycoplasma pneumoniae [1]. Antibiotics are not indicated for the initial treatment of acute bronchitis and when prescribed can do more harm than good. In 2014, 266.1 million courses of antibiotics were dispensed to outpatients in U.S. community pharmacies with at least 30 percent of those antibiotics being potentially unnecessary prescriptions [2].

A 2017 Cochrane review of 17 studies assessing outcomes and adverse effects of antibiotic use in children and adults with acute bronchitis found limited evidence of clinical benefit to support the use of antibiotics across all age ranges studied. For eleven studies at follow-up, there was no difference in participants described as being clinically improved between the antibiotic and placebo groups. Additionally, the review found a small but significant increase in adverse effects in people treated with antibiotics. The most common side effects included nausea, vomiting, diarrhea, headache and rash [3]. Guidelines recommend against the use of antibiotics in patients [3, 4, 5].

References:

[1] Hart, A.M. 2014. "Evidence-Based Diagnosis and Management of Acute Bronchitis." Nurse Practitioner. 39(9):32-39. Doi: 10.1097/01.NPR.0000452978.99676.2b.

[2] Centers for Disease Control and Prevention (CDC). 2017. Antibiotic Prescribing and Use in Doctor's Offices. What is Acute Bronchitis? https://www.cdc.gov/antibiotic-use/community/for-patients/common-illnesses/bronchitis.html

[3] Smith, S.M., T., Fahey, T., Smucny, J., Becker, L.A. 2017. "Antibiotics for Acute Bronchitis." Cochrane Database Syst Rev DOI: 10.1002/14651858.CD000245.pub4

[4] Kinkade, S. & Long, N. A. (2016). Acute Bronchitis. American Academy of Family Physicians, 94(7), 560-565.

[5] Ralston, S. L., Leiberthal, A. S., Meissner, H. C., Alverson, B. K., Baley, J. E., Gadomski, A. M., et al. (2014). Clinical Practice Guideline: The Diagnosis, Management, and Prevention of Bronchiolitis. Pediatrics, 134, e1474-e1502.

S.4. Numerator Statement: The number of dispensed antibiotic medications following an episode of acute bronchitis/bronchiolitis. The measure is reported as an inverted rate (i.e., 1 – numerator/denominator) to reflect the proportion of episodes during which an antibiotic was not dispensed (a higher rate is better).

S.6. Denominator Statement: Episodes for members aged 3 months and older with a diagnosis of acute bronchitis or bronchiolitis during the intake period.

S.8. Denominator Exclusions: As listed in the denominator details, the final denominator population does not include episodes with a history of select comorbid conditions, history of antibiotic use, or presence of a competing diagnosis

De.1. Measure Type: Process

S.17. Data Source: Claims

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Jan 07, 2013

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2012

- This is a maintenance process measure using health plan claims data to assess the percentage of episodes for members ages 3 months and older with a diagnosis of acute bronchitis/bronchiolitis that did not result in an antibiotic dispensing event.
- Developer provides a logic model depicting a reduction in the inappropriate dispensing of antibiotics for acute bronchitis/bronchiolitis would lead to fewer strains of antibiotic-resistant pathogens and a reduction in community-acquired antibiotic resistant infections.
- In its 2012 submission, the developer cited Cochrane systematic review and a literature review which both found limited evidence for antimicrobials for acute bronchitis.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- The developer cited a 2016 clinical practice guideline for Acute Bronchitis from the American Academy of Family Physicians which recommended that clinicians, "Avoid prescribing antibiotics for uncomplicated acute bronchitis. (Grade A Recommendation)
- The developer cited a 2017 Cochrane Review for antibiotics for acute bronchitis with no grade assignment that found, limited evidence of clinical benefit to support the use of antibiotics in acute bronchitis. The developer states that no specific grade was assigned for the trials cited in the review however, the quality of the trials was generally good.
- The developer cites a 2014 clinical practice guideline for the diagnosis, management, and prevention of bronchiolitis from the American Academy of Pediatrics which recommended that, clinicians should not administer antibacterial medications to infants and children with a diagnosis of bronchiolitis unless there is a concomitant bacterial infection, or a strong suspicion of one. (Grade B Recommendation)

Exception to evidence

Not Applicable

Questions for the Committee:

- What is the relationship of this measure to patient outcomes?
- How strong is the evidence for this relationship?
- Is the evidence directly applicable to the process of care being measured?

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency: high (Box 5) → High

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- For the current submission, the developer provided the following commercial and Medicaid rates:
    - For 2017 through 2019, the commercial HEDIS data below covers 406, 387, and 393 health plans respectively.
        - 2019: Mean: 40.8%; Min: 10.8%;10th - 90th Percentile Range: 19.3%-53.8%; Max: 86.5%; IQR: 10.9%

        - 2018: Mean: 33.7%; Min: 10.9%;10th - 90th Percentile Range: 14.8%-48.6%; Max: 81.5%; IQR: 11.8%
        - 2017: Mean: 30.8%; Min: 10.7%;10th - 90th Percentile Range: 13.4%-43.9%; Max: 80.5%; IQR: 9.9%
    - For 2017 through 2019, the Medicaid HEDIS data below covers 213, 213, and 235 health plans respectively.
        - 2019: Mean: 52.2%; Min: 11.1%; 10th - 90th Percentile Range: 28.8%-65.2%; Max: 100.0%; IQR: 13.2%
        - 2018: Mean: 36.3%; Min: 9.6%; 10th - 90th Percentile Range:19.5%-48.9%; Max: 80.4%; IQR: 11.2%
        - 2017: Mean: 33.7%; Min: 9.5%; 10th - 90th Percentile Range:12.1%-44.6%; Max: 76.5%; IQR: 9.7%

Disparities

- The developer noted that HEDIS data are stratified by type of insurance (e.g., Commercial, Medicaid, Medicare), but could be stratified by demographic variables.
- The developer noted that HEDIS does include two measures which can be used to assess disparities in the health plan population.
- The developer summarized literature addressing disparities and acute bronchitis. One study found that patients who are white or have commercial insurance were more likely to inappropriately receive antibiotics for acute bronchitis while another study found that racial and ethnic minorities were less likely to receive antibiotics when they are appropriate to prescribe.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?

Preliminary rating for opportunity for improvement: ☒ High ☐ Moderate ☐ Low ☐ Insufficient

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- Evidence supports measure
- This is a process measure examining the use of antibiotics for acute bronchitis/bronchiolitis. There is adequate evidence to support this measure. I am not aware of any new evidence.
- Evidence is updated and applies directly to process being measured.
- Evidence to support measure focus is sound
- Good evidence that most episodes of bronchitis are viral and that treatment without antibiotics leads to same outcomes as with. Cites 2017 Cochrane review of 17 studies to support position
- Evidence is strong for the question.
- This is a maintenance direct process measure. The desired outcome is fewer prescriptions written to treat acute bronchitis/bronchiolitis. The update to the literature review on the cost/benefit analysis of

such treatment continues to argue against antibiotics for most adults and children diagnosed with acute bronchitis/bronchiolitis.

- Evidence is strong for this measure. Supported by the AAFP and the AAP in clinical practice guidelines. A 2017 review of evidence found limited clinical evidence to support the use of antibiotics for acute bronchitis.
- Process measure, evidence rating is high

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- Gap exists which provides opportunity for improvement
- There have been some changes in the measurement specification to expand the age range; however, the data demonstrates a less than optimal performance. There is no data on subgroups provided.
- Performance data show improvement but still less than optimal performance.
- yes, improving performance gap year to year, but there is still a gap
- Disparities not able to be collected within the measure data collection. Review of antibiotic use for bronchitis reveals differences in that whites were more likely to receive antibiotics when perhaps not appropriate, and other populations were less likely to receive antibiotics when perhaps appropriate. There continues to be overuse of antibiotics in all subgroups tested without significant change over the past 3 years.
- Performance gap is worth monitoring to see if improvement for reducing inappropriate antibiotics and improving use of appropriate even for uninsured or lower income.
- Claims data were analyzed from health plans for years 2017 to 2019. The health plans were divided into two groups. Both groups showed an increase in the percentage of encounters with the diagnosis of acute bronchitis/bronchiolitis in which prescriptions were NOT written, with an increase in both groups over the three years of data analysis with the best score at 52%. The optimal percentage is not cited, but is presumed to be significantly closer to 100% than 50%. HEDIS measurements are used and do not allow analysis by social and economic categories. There was a literature review for studies investigating disparities in antibiotic prescribing for acute bronchitis by race.
- Evidence provided by the measure steward indicated a performance gap/disparity in white versus black populations. A disparity or gap was also identified in commercial populations compared to Medicaid populations.
- Yes, a performance gaps exists

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

2c. For composite measures: empirical analysis support composite approach

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or

that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**

Evaluators: NQF Staff
Staff Review

Specifications:

- Since last endorsement, the developer expanded the eligible population by broadening the age range and including the Medicare line of business as well changing the measure to an episode-based measure.

Reliability

- The developer conducted measure score level reliability testing.
    - Using 2019 HEDIS data, the developer used a beta-binominal model to assess the signal-to-noise ratio. Using this method, the mean commercial reliability score was 0.963 and the mean Medicaid reliability score was 0.982.
    - For signal to noise, the developer states a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. The testing suggests that all indicators within this measure have good reliability between 0.7 and 1.0.

Validity

- Validity testing was performed at the measure score level through construct validity testing.
    - The developer conducted Pearson correlation for construct validity using HEDIS health plan data from two measures.
        - Results:
            - Positive Correlation: Appropriate Treatment for Upper Respiratory Infection
                - Medicaid: Correlation coefficient = 0.68, $p < 0.001$
                - Commercial: Correlation coefficient = 0.68, $p < 0.001$
            - Negative Correlation: Antibiotic Utilization
                - Medicaid: Correlation coefficient = -0.60, $p < 0.001$
                - Commercial: Correlation coefficient = -0.64, $p < 0.001$
        - The developer concluded that plans that perform well on this measure are likely to perform well on the positively correlated measures.
- The developer also conducted face validity for this measure. Provided face validity does not meet current NQF requirements, however, since empirical validity testing has been conducted, face validity is not needed.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?

Preliminary rating for reliability:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**
Preliminary rating for validity:	☒ **High**	☐ **Moderate**	☐ **Low**	☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- The number of dispensed antibiotic medications - can member be dispensed antibiotics more than once per episode?
- No concerns about reliability specifications.
- Data elements are clear. No concerns.
- reliability - specifications are appropriate
- Reliability was high in the populations tested. To improve performance in health systems, the focus on outpatient visits, plus observation and ED events perhaps poses challenges for practice improvement opportunities. This, however, does not impede consistent implementation.
- I have no specific concerns for reliability.
- Steps are clear and well described and would support consistent implementation. The denominator was enlarged from the original application by including plan members of age 3 months or older.
- Data elements are clearly defined. The measure steward conducted empirical validity testing which results indicated strong reliability of the measure specifications.
- Reliability rating is high

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- measure score level reliability testing
- None.
- No concerns.
- no
- no
- It appears that the measure stands up well in all comparisons.
- Reliability was analyzed by the beta binomial method (signal to noise reliability). Results using this method are consistently better than 0.7, the minimal threshold for reliability.
- No concerns.
- No concerns

2b1. Validity -Testing: Do you have any concerns with the testing results?

- empirical validity testing provided
- no
- No concerns

- no
- None
- It appears the to be very valid.
- Construct validity was tested using Pearson correlations on two HEDIS measures - appropriate treatment for upper respiratory infections and average antibiotic prescriptions per member per year. No significant concerns with these results.
- No validity testing concerns.
- No concerns

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment) 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- No Risk adjustment
- no risk adjustment performed
- Exclusions are consistent with evidence.
- no major concerns
- Exclusions are for proximal use of antibiotics for other diagnoses. There is no risk adjustment or stratification.
- At present, I see no issues with exclusions or risk adjustment.
- No risk adjustment or stratification
- No concerns with the exclusion criteria.
- No concerns

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- na
- no
- No threats to validity.
- no
- Not likely. Based upon claims data. A small percentage of individuals potentially could pay cash for antibiotics and thus not be counted as part of the measure
- No issues with threats to validity.
- Meaningful Differences - Tested by inter-quartile range tests that were statistically significant. Comparability of performance scores - There is only one set of specifications. Missing data - NCQA auditors did not find sources of missing data for this measure. Exclusions were reasonable and limited.
- Reliable data sources are used for this measure. No concerns about the ability to have complete data for measurement purposes. The measure specifications should result in the ability to identify meaningful differences for quality improvement purposes. Results should be comparable.

- No concerns

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data is collected during the provision of care and coded by someone other than person obtaining original information.
- All data elements are in defined fields in electronic claims.
- The developer conducts audits for all HEDIS collection and reporting processes. An independent audit of HEDIS process to verify integrity of HEDIS collection and reporting system is conducted and a Policy Clarification Support System is used to generate ongoing feedback from measure users.
- The developer notes that noncommercial uses do not require the consent of the measure developer. However, commercial use of the measure requires the prior written consent of NCQA.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

1. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?

- data elements are in defined fields in electronic claims
- no concerns. all elements should be easily obtainable.
- Data elements are routinely generated during care delivery. No concerns
- no major concerns
- Electronic capture from billing. High Feasibility
- Has been used and little issue foreseen for future.
- The data elements are routinely generated and available from the electronic forms. The process is subject to the HEDIS Compliance Audit. The data collection process for this measure has been operating for three years.
- No concerns with feasibility. All data elements should be available to calculate the measure. All data elements are Data generated during care delivery. Data elements should be routinely available in an administrative form. No concerns with the data collection strategy.
- Process measure based on claims, information routinely generated in the delivery of healthcare

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No** ☐ **UNCLEAR**

Accountability program details

- The measure is part of the following programs: NCQA Quality Compass; NCQA Health Plan Rating/Report Cards; NCQA Health Plan Accreditation; Integrated Healthcare Association; CDC Measuring Outpatient Antibiotic Prescribing; CDC Core Elements Of Outpatient Antibiotic Stewardship

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- NCQA publishes HEDIS results annually and presents at various conferences and webinars.
- Technical assistance is provided on measures through the developer's Policy Clarification Support System.
- NCQA utilizes a consensus-based process to obtain broad input on the measure from several multi-stakeholder advisory panels, public comment posting, and questions submitted to the Policy Clarification Support System.
- The developer notes that they have not received any feedback which would indicate any barriers to measure implementation.

Questions for the Committee:

- How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

a. The developer notes that improvement results cannot be reported for this measure due to the change in measure denominator age ranges from 2018 to 2019.

b. The developer notes that for 2019, the average performance rate was 40.8% for commercial plans and 33.7% for Medicaid plans.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

c. The developer notes that the intended benefits of this measure outweigh the potential harms.

d. The developer does not list any potential harms.

Potential harms

e. The developer states there were no identified potential harms for this measure during testing or since implementation.

***Questions for the Committee**:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

Preliminary rating for Usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? 4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- measure is part of multiple programs
- results published annually; technical assistance is provided; input is obtained from advisory panels, public comment and through policy clarification support system. No feedback received indicating barriers to implementation.
- Publicly reported and used in accountability programs listed in application
- no major concerns
- Used as part of performance reporting without feedback from those being provided feedback. Challenge for identifying feedback to providers for where in the health system appropriate/inappropriate prescribing may have occurred. (PCP, ED, etc.)
- Meets the requirements for Use.
- The measure has been reported through NCQA and CDC publications. it is used for NCQA health plans accreditation and quality improvement projects. There are ongoing processes to solicit and evaluate feedback on this measure from stakeholders and the public.

- The measure is being used by NCQA (the measure steward) in Quality Compass; health plan rating cards; health plan accreditation; and by the CDC for two programs. No concerns with the measure meeting use criteria for endorsement.
- Currently in use in multiple programs

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- improvement not reported for this measure due to the change in measure denominator
- none
- No harms
- no concerns
- Although the developer does not list harms or unintended consequences, with the increased emphasis on value based payments, and pressures placed upon providers to meet measures, will there be a mindset of strict avoidance of antibiotic use leading to those that do need treatment being deferred? Why is there a gap in prescribing to whites and non-whites? Nonetheless, this is an important measure for improving health outcomes, and reducing care expenditures
- Appears to meet the requirements for Usability.
- Establishing performance in improvement is complicated by the change in the denominator in 2019 (including ages down to 3 months old). There are no identified unintended consequences with this measure.
- The measure has been used in performance measurement programs although comparability is not appropriate due to changes in measure age ranges 2018 - 2019 and expanding to include Medicare line of business. Also now an episode-based measure. No identified harms from use of the measure. No identified unintended consequences.
- No concerns

Criterion 5: Related and Competing Measures

Related or competing measures

f. 0069: Appropriate Treatment for Upper Respiratory Infection

Harmonization

g. The developer indicated that this measure has been harmonized with NQF 0069.

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?
   - no additional steps apparent.
   - Harmonized with competing measure NQF0069
   - appropriately harmonized with 0069
   - 0069 - developer states this measure has been harmonized.
   - Harmonized with other related measures.

- 0069 Appropriate Treatment for Upper Respiratory Infection is listed as related or competing. The developer noted that 0069 has been harmonized with this measure (but no details or explanation were provided).
- Measure was harmonized with NQF 0069: Appropriate Treatment for Upper Respiratory Infections.
- No concerns

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/21/2021

- No NQF Members have submitted support/non-support choices as of this date.
- No Public or NQF Member comments submitted as of this date.

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0058

Measure Title: Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (AAB)

Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**
☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☐ **Facility** ☒ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☐ **New** ☒ **Previously endorsed** (NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   Submission document: "MIF_0058" document, items S.1-S.22

   ***NOTE**: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**
   - Since last endorsement, the developer expanded the eligible population by broadening the age range and including the Medicare line of business as well changing the measure to an episode-based measure.
   - No concerns

RELIABILITY: TESTING

Submission document: "MIF_0058" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of **patient-level data** conducted?

   ☐ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   - The developer used a beta-binominal model to assess the signal-to-noise ratio of the measure score.
   - The reliability of the measure score was assessed using 2019 HEDIS data.
   - A minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities.
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   - The developer provided mean signal-to-noise reliability for commercial and Medicaid plans for all ages as well as ages 3 months to 17 years; age 18-64; and, 65 years and older.

Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis	**Point estimate: Mean Signal-To-Noise Reliability** (Commercial)	**Point estimate: Mean Signal-To-Noise Reliability** (Medicaid)
Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (Total)	0.9634	0.9815
Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (age 3 Months-17 Years)	0.9216	0.9829
Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (age 18-64)	0.9539	0.9295
Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (age 65+)	0.8861	0.9609

   - The developer states that the reliability estimates provided indicate good reliability.
8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (data element testing was not performed)
10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and **all** testing results):

☒ **High** (NOTE: Can be HIGH **only if** score-level testing has been conducted)

☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has **not** been conducted)

☐ **Low** (NOTE: Should rate **LOW** if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: Should rate **INSUFFICIENT** if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - A suitable method was used for reliability testing and scores indicate good reliability for all product lines and for each age group.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

    Submission document: Testing attachment, section 2b2.
    - This measure does not include: episodes with a history of select comorbid conditions, patients with a history of antibiotic use, or presence of a competing diagnosis.
    - No concerns.
13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    Submission document: Testing attachment, section 2b4.
    - To demonstrate meaningful differences in performance, the developer calculated an inter-quartile range for both commercial and Medicaid health plans.
    - For commercial plans, the IQR was 11% and for Medicaid plans, the IQR was 13%.
    - No concerns.
14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Testing attachment, section 2b5.
    - Not Applicable
15. **Please describe any concerns you have regarding missing data.**

    Submission document: Testing attachment, section 2b6.
    - The developer indicated that it audits the diagnostic and procedure code fields for this measure.
    - The developer reported that no missing data was found during the audit.
    - No concerns.
16. **Risk Adjustment**

    16a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

    16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

    ☐ Yes ☐ No ☒ Not applicable

    16c. **Social risk adjustment:**

    16c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☐ Not applicable

    16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

16d. **Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

16d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

16e**. Assess the risk-adjustment approach**

Not Applicable

For cost/resource use measures ONLY:

17. **Are the specifications in alignment with the stated measure intent?**

☐ Yes ☐ Somewhat ☐ No (If "Somewhat" or "No", please explain)

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

19. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**

20. **Method of establishing validity of the measure score:**

☐ Face validity

☒ Empirical validity testing of the measure score

☐ N/A (score-level testing not conducted)

21. **Assess the method(s) for establishing validity**

Submission document: Testing attachment, section 2b2.2

- The developer assessed construct validity on 2019 HEDIS data by calculating Pearson Correlation Coefficient between this measure and the HEDIS measure Appropriate Treatment for Upper Respiratory Infection; hypothesizing a positive correlation.
- The developer also compared this measure to the HEDIS measure Antibiotic Utilization; hypothesizing a negative correlation.
- The developer stated that face validity was provided but it did not meet NQF requirements. However, since empirical validity testing was provided, face validity is not needed.

22. **Assess the results(s) for establishing validity**

Submission document: Testing attachment, section 2b2.3

- The developer found a positive correlation between this measure and Appropriate Treatment for Upper Respiratory Infection.
    - Medicaid: Correlation coefficient = 0.68, $p < 0.001$
    - Commercial: Correlation coefficient = 0.68, $p < 0.001$
- The developer found a negative correlation between this measure and Antibiotic Utilization.
    - Medicaid: Correlation coefficient = -0.60, $p < 0.001$
    - Commercial: Correlation coefficient = -0.64, $p < 0.001$

- The developer concluded that plans that perform well on this measure are likely to perform well on the positively correlated measures.

23. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (score-level testing was not performed)

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**
    NOTE that data element validation from the literature is acceptable.

    Submission document*: Testing attachment, section 2b1.*

    ☐ **Yes**

    ☐ **No**

    ☒ **Not applicable** (data element testing was not performed)

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe that there **are** threats to validity and/or relevant threats to validity were **not assessed OR** if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level **is required**; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**
    - Score level testing was conducted.
    - Correlation analysis demonstrated construct validity of this measure.
    - The developer provided data related to exclusions and meaningful differences. The developer noted that they did not find any issues with missing data.
    - In the previous review, the committee expressed concerns about validity due to the potential shift in diagnosis because it reflects one billing code; a simple change to "bronchitis not specified" will miss the cases.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

    - No additional concerns or questions.

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

AAB_0058_Evidence_Form-637400848107413979.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

1a. Evidence (subcriterion 1a)

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Outcome:

☐Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)

☐ Intermediate clinical outcome (*e.g., lab value*):

☒ Process:

☒ Appropriate use measure: appropriate antibiotic treatment for acute bronchitis/bronchiolitis

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Reduction in the inappropriate dispensing of antibiotics for acute bronchitis/bronchiolitis → fewer strains of antibiotic-resistant pathogens → reduction in community-acquired antibiotic resistant infections.

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

N/A

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ **Clinical Practice Guideline recommendation** (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ **Other systematic review and grading of the body of evidence** (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Table 1. American Academy of Family Physicians Clinical Practice Guidelines, 2016

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	• American Academy of Family Physicians • Acute Bronchitis • Kinkade & Long • 2016 • Kinkade, S. & Long, N. A. (2016). Acute Bronchitis. *American Academy of Family Physicians, 94*(7), 560-565. • https://www.aafp.org/afp/2016/1001/p560.html#ref-list-1
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Avoid prescribing antibiotics for uncomplicated acute bronchitis.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	*
Provide all other grades and definitions from the evidence grading system	*
Grade assigned to the **recommendation** with definition of the grade	A – Recommendation based on consistent and good-quality patient-oriented evidence.

Systematic Review	Evidence
Provide all other grades and definitions from the recommendation grading system	B – Recommendation based on inconsistent or limited-quality patient-oriented evidence. C – Recommendation based on consensus, usual practice, opinion, disease-oriented evidence, or case series for studies of diagnosis, treatment, prevention, or screening.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	There were 48 studies and references cited for this recommendation. The studies referenced were mostly randomized-controlled trials and systematic reviews, study types that are considered to be of good quality.
Estimates of benefit and consistency across studies	Benefit of antibiotic avoidance: - Decrease in antimicrobial resistant infections. - Decrease in adverse effects of antibiotics, such as nausea, vomiting and allergic reactions. Consistency was not directly assessed.
What harms were identified?	*
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Conclusions have not changed.

*cell intentionally left blank

Table 2. Cochrane Database Systematic Review, 2017

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	Cochrane Database of Systematic Reviews • Antibiotics for Acute Bronchitis (Review) • Smith, Fahey, Smucny, & Becker • 2017 • Smith, S. M., Fahey, T., Smucny, J., & Becker, L.A. (2017). Antibiotics for Acute Bronchitis (Review). *Cochrane Database of Systematic Reviews*, Issue 6. Art. No.: CD000245. Doi:10.1002/14651858.CD000245.pug4. • https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6481481/pdf/CD000245.pdf

Systematic Review	Evidence
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	There is limited evidence of clinical benefit to support the use of antibiotics in acute bronchitis. Antibiotics may have a modest beneficial effect in some patients such as frail, elderly people with multimorbidity who may not have been included in trials to date. However, the magnitude of this benefit needs to be considered in the broader context of potential side effects, medicalization for a self-limiting condition, increased resistance to respiratory pathogens, and cost of antibiotic treatment.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The quality of these trials was generally good, particularly for more recent studies. No specific grade was assigned.
Provide all other grades and definitions from the evidence grading system	*
Grade assigned to the **recommendation** with definition of the grade	*
Provide all other grades and definitions from the recommendation grading system	*
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	This was an update to an original evidence review by Cochrane. 9 trials involving over 750 patients aged eight to over 65 and including smokers and non-smokers were included Cochrane Review. The 9 studies were randomized controlled trials comparing any antibiotic therapy with placebo in acute bronchitis or acute productive cough without other obvious cause in patients without underlying pulmonary disease. Original review: Smucny J, Fahey T, Becker L, Glazier R, McIsaac W. Antibiotics for acute bronchitis (Cochrane Review). Cochrane Database Syst Rev 2000;4: CD000245.
Estimates of benefit and consistency across studies	- Decrease in medication adverse effects like nausea and vomiting. - Lower incidence of antibiotic resistance downstream. - Lower healthcare spending and utilization for complications and hospitalizations due to antibiotic resistance. Consistency was not directly assessed.

Systematic Review	Evidence
What harms were identified?	No harms were identified from avoiding antibiotic use for bronchitis. However, potential harms of using antibiotics for bronchitis were noted, including antibiotic resistance and side effects.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	This is the most recent review of the evidence for the use of antibiotics for acute bronchitis. Between the original review and the current review, the conclusions did not change.

*cell intentionally left blank

Table 3. American Academy of Pediatrics, 2014

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	American Academy of Pediatrics • Clinical Practice Guideline: The Diagnosis, Management, and Prevention of Bronchiolitis • Shawn L Ralston, MD, et al. • 2014 • Ralston, S. L., Leiberthal, A. S., Meissner, H. C., Alverson, B. K., Baley, J. E., Gadomski, A. M., et al. (2014). Clinical Practice Guideline: The Diagnosis, Management, and Prevention of Bronchiolitis. *Pediatrics, 134*, e1474-e1502. • https://pediatrics.aappublications.org/content/pediatrics/134/5/e1474.full.pdf
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Clinicians should not administer antibacterial medications to infants and children with a diagnosis of bronchiolitis unless there is a concomitant bacterial infection, or a strong suspicion of one.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	B – Trials or diagnostic studies with minor limitations; consistent findings from multiple observational studies

Systematic Review	Evidence
Provide all other grades and definitions from the evidence grading system	A – Well designed and conducted intervention trials; meta analyses on applicable populations; independent gold standard studies of applicable populations C – Single or few observational studies or multiple studies with inconsistent findings or major limitations D – Expert opinion, case reports, reasoning from first principles
Grade assigned to the **recommendation** with definition of the grade	Strong – Action is favored because anticipated benefits clearly exceed harms (or vice versa), and quality of evidence is excellent or unobtainable
Provide all other grades and definitions from the recommendation grading system	Moderate – Action is favored because anticipated benefits clearly exceed harms and the quality of evidence is good but not excellent Weak – Action is favored because anticipated benefits clearly exceed harms but the quality of evidence is weak
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	There were 28 studies and references cited for this recommendation. The studies referenced were mostly randomized-controlled trials and systematic reviews, study types that are considered to be of good quality.
Estimates of benefit and consistency across studies	Authors highlighted several randomized controlled trials that showed there was no benefit from using antibiotics for bronchiolitis. Consistency was not directly assessed.
What harms were identified?	*
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	This was the most recent review.

*cell intentionally left blank

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

The vast majority of acute bronchitis cases are viral. Bacteria are detected in 1% to 10% of cases, and can include Bordetella pertussis, Chlamydia pneumoniae, and Mycoplasma pneumoniae [1]. Antibiotics are not indicated for the initial treatment of acute bronchitis and when prescribed can do more harm than good. In 2014, 266.1 million courses of antibiotics were dispensed to outpatients in U.S. community pharmacies with at least 30 percent of those antibiotics being potentially unnecessary prescriptions [2].

A 2017 Cochrane review of 17 studies assessing outcomes and adverse effects of antibiotic use in children and adults with acute bronchitis found limited evidence of clinical benefit to support the use of antibiotics across all age ranges studied. For eleven studies at follow-up, there was no difference in participants described as being clinically improved between the antibiotic and placebo groups. Additionally, the review found a small but significant increase in adverse effects in people treated with antibiotics. The most common side effects included nausea, vomiting, diarrhea, headache and rash [3]. Guidelines recommend against the use of antibiotics in patients [3, 4, 5].

References:

[1] Hart, A.M. 2014. “Evidence-Based Diagnosis and Management of Acute Bronchitis.” Nurse Practitioner. 39(9):32-39. Doi: 10.1097/01.NPR.0000452978.99676.2b.

[2] Centers for Disease Control and Prevention (CDC). 2017. Antibiotic Prescribing and Use in Doctor’s Offices. What is Acute Bronchitis? https://www.cdc.gov/antibiotic-use/community/for-patients/common-illnesses/bronchitis.html

[3] Smith, S.M., T., Fahey, T., Smucny, J., Becker, L.A. 2017. “Antibiotics for Acute Bronchitis.” Cochrane Database Syst Rev DOI: 10.1002/14651858.CD000245.pub4

[4] Kinkade, S. & Long, N. A. (2016). Acute Bronchitis. American Academy of Family Physicians, 94(7), 560-565.

[5] Ralston, S. L., Leiberthal, A. S., Meissner, H. C., Alverson, B. K., Baley, J. E., Gadomski, A. M., et al. (2014). Clinical Practice Guideline: The Diagnosis, Management, and Prevention of Bronchiolitis. Pediatrics, 134, e1474-e1502.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

The following HEDIS data reflect the most recent years of measurement for this measure. It includes number of health plans, percentiles, mean, min, max and standard deviations.

Data are summarized at the health plan level (i.e. “N” represents the number of health plans). All rates are reported as an inverted rate (i.e. 1- numerator/denominator). The rates for MY 2017 and MY 2018 reflect the percentage of health plan members that were not dispensed an antibiotic (18-64 years). For MY 2019, the rate reflects the percentage of episodes that were not dispensed antibiotic (3 months and older).

Data are stratified by year and product line (i.e. commercial, Medicaid). Medicare was specified for the measure starting MY 2019, but CMS did not require health plans to report for MY 2019 due to the COVID-19 pandemic; therefore, Medicare performance rates are not provided.

Commercial

YEAR|N|MEAN|STDEV|MIN|10th|25th|50th|75th|90th|MAX|Interquartile Range

2019|406|40.8%|10.8%|19.3%|29.8%|33.8%|39.3%|44.7%|53.8%|86.5%|10.9%

2018|387|33.7%|10.9%|14.8%|23.2%|26.4%|31.5%|38.1%|48.6%|81.5%|11.8%

2017|393|30.8%|10.7%|13.4%|21.2%|24.2%|28.4%|34.1%|43.9%|80.5%|9.9%

Medicaid

YEAR|N|MEAN|STDEV|MIN|10th|25th|50th|75th|90th|MAX|Interquartile Range

2019|213|52.2%|11.1%|28.8%|39.9%|45.0%|50.7%|58.1%|65.2%|100.0%|13.2%

2018|213|36.3%|9.6%|19.5%|27.0%|29.8%|34.1%|41.1%|48.9%|80.4%|11.2%

2017|235|33.7%|9.5%|12.1%|25.2%|27.6%|32.0%|37.4%|44.6%|76.5%|9.7%

Note: Data from 2017 and 2018 shows performance on the AAB measure before revisions (i.e., change to episode-based denominator, etc.) were made. Data from 2019 shows performance on the revised measure.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

HEDIS data are stratified by type of insurance (e.g., Commercial, Medicaid, Medicare). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities if the data are available to a plan. NCQA is actively engaged with partners including the CMS Office of Minority Health in identifying feasible methods to further integrate social risk factors into health plan quality measures, with a focus on stratification. Our work is aligned with recent recommendations from MedPAC and ASPE on optimal methods for addressing social risk in quality measurement and programs.1,2 This is an NCQA wide initiative. Our intent is to implement methods to bridge data concerns in the future.

HEDIS includes two measures that can be used as tools for assessing race/ethnicity and language needs of a plan's population: Race/Ethnicity Diversity of Membership and the Language Diversity of Membership. These measures promote standardized methods for collecting these data and follow Office of Management and Budget and National Academy of Medicine guidance for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing, and using race/ethnicity and language data to assess health care disparities.

1. Medicare Payment Advisory Commission. (2020). The Medicare Advantage program: Status report. In Report to the Congress: Medicare Payment Policy (p. 397). http://medpac.gov/docs/default-source/reports/mar20_medpac_ch13_sec.pdf

2. Office of the Assistant Secretary for Planning and Evaluation, & U.S. Department of Health & Human Services. (2020). Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-basedpurchasing-programs

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

Demographic and socioeconomic factors can affect antibiotic prescribing. A 2018 study of 448,990 outpatient visits for common upper respiratory conditions, such as bronchitis, that should not require antibiotics found that adult patients who were white or had commercial insurance were significantly more likely to receive inappropriate antibiotic treatment. Additional factors that increased the likelihood of receiving antibiotic treatment included provider type, age of the provider and practice setting [6].

Studies to determine whether racial and ethnic differences exist in antibiotic prescribing among children in the U.S. have found that, when compared to white children, black and other racial and ethnic minorities are less likely to receive antibiotics for acute respiratory tract infections. A 2009 study of 1,296,517 encounters by over 200,000 children to 222 clinicians in 25 practices found that when treated by the same clinician, black children received fewer antibiotic prescriptions, fewer acute respiratory tract infection diagnoses and a lower proportion of broad-spectrum antibiotic prescriptions than nonblack children [7]. A 2017 study of 39,445 pediatric emergency department encounters for viral acute respiratory tract infections found that 4.3 percent of white children received antibiotics, compared to just 1.9 percent of black, 2.6 percent of Hispanic and 2.9 percent of other Non-Hispanic children. Factors such as parental expectations, provider perceptions of parental expectations and implicit provider biases may contribute to the racial and ethnic differences in overprescribing [8].

[6] Schmidt, M.L., M.D. Spencer, L.E. Davidson. 2018. "Patient, Provider, and Practice Characteristics Associated with Inappropriate Antimicrobial Prescribing in Ambulatory Practices." Infection Control and Hospital Epidemiology. 39(3): 307-315. doi: 10.1017/ice.2017.263.

[7] Gerber, J.S., P.A. Praad, A.R. Localio, A.G. Fiks, et al. 2013. "Racial Differences in Antibiotic Prescribing by Primary Care Physicians." Pediatrics. 131(4):677-684. doi:10.1542/peds.2012-2500.

[8] Goyal, M.K., T.J. Johnson, J.M. Chamberlain, C. Casper, et al. 2017. "Racial and Ethnic Differences in Antibiotic Use for Viral Illness in Emergency Departments." Pediatrics.140(2):e20170203. doi: https://doi.org/10.1542/peds.2017-0203.

Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, **as specified**, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

Infectious Diseases (ID), Infectious Diseases (ID) : Pneumonia and respiratory infections

De.6. Non-Condition Specific*(check all the areas that apply):*

Safety : Overuse

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

N/A

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment**:** 0058_AAB_Fall_2020_Value_Sets.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

This measure recently underwent several changes, which are detailed below:

- Expanded the eligible population by broadening the age range and including the Medicare line of business.
  - Rationale: Clinical guidelines recommend against the use of antibiotics to treat patients diagnosed with bronchitis regardless of age. To broaden the coverage of the measure, the age range was adjusted from members 18-64 years of age to those 3 months or older. The Medicare line of business was tested and added.
- Changed to an episode-based measure.
  - Rationale: The member-based denominator resulted in members with multiple bronchitis diagnoses throughout the measurement period counting once. An episode-based measure captures more episodes of potentially inappropriate antibiotic treatment and measure testing indicated that an episode-based measure increased denominator sizes by capturing more treatment episodes but had a small impact on performance rates.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The number of dispensed antibiotic medications following an episode of acute bronchitis/bronchiolitis. The measure is reported as an inverted rate (i.e., 1 – numerator/denominator) to reflect the proportion of episodes during which an antibiotic was not dispensed (a higher rate is better).

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Dispensed prescription for an antibiotic medication (listed in Table AAB Antibiotic Medications) on or three days after the episode date.

Table AAB Antibiotic Medications

Aminoglycosides: Amikacin; Gentamicin; Streptomycin; Tobramycin

Aminopenicillins: Amoxicillin; Ampicillin

Beta-lactamase inhibitors: Amoxicillin-clavulanate; Ampicillin-sulbactam; Piperacillin-tazobactam; Ticarcillin-clavulanate

First-generation cephalosporins: Cefadroxil; Cefazolin; Cephalexin

Fourth-generation cephalosporins: Cefepime

Ketolides: Telithromycin

Lincomycin derivatives: Clindamycin; Lincomycin

Macrolides: Azithromycin; Clarithromycin; Erythromycin; Erythromycin ethylsuccinate; Erythromycin lactobionate; Erythromycin stearate

Miscellaneous antibiotics: Aztreonam; Chloramphenicol; Dalfopristin-quinupristin; Daptomycin; Erythromycin-sulfisoxazole; Linezolid; Metronidazole; Vancomycin

Natural penicillins: Penicillin G benzathine-procaine; Penicillin G potassium; Penicillin G procaine; Penicillin G sodium; Penicillin V potassium; Penicillin G benzathine

Penicillinase resistant penicillins: Dicloxacillin; Nafcillin; Oxacillin

Quinolones: Ciprofloxacin; Gemifloxacin; Levofloxacin; Moxifloxacin; Norfloxacin; Ofloxacin;

Rifamycin derivatives: Rifampin

Second generation cephalosporin: Cefaclor; Cefotetan; Cefoxitin; Cefprozil; Cefuroxime

Sulfonamides: Sulfadiazine;; Sulfamethoxazole-trimethoprim

Tetracyclines: Doxycycline; Minocycline; Tetracycline

Third generation cephalosporins: Cefdinir; Cefditoren; Cefixime; Cefotaxime; Cefpodoxime; Ceftazidime; Ceftibuten; Ceftriaxone

Urinary anti-infectives: Fosfomycin; Nitrofurantoin; Nitrofurantoin macrocrystals-monohydrate; Trimethoprim; Nitrofurantoin macrocrystals

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

Episodes for members aged 3 months and older with a diagnosis of acute bronchitis or bronchiolitis during the intake period.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Members who had an outpatient visit (Outpatient Value Set), a telephone visit (Telephone Visits Value Set), an e-visit or virtual check-in (Online Assessments Value Set), an observation visit (Observation Value Set) or an ED visit (ED Value Set) during the Intake Period, with a diagnosis of acute bronchitis/bronchiolitis (Acute Bronchitis Value Set).

Do not include visits that result in an inpatient stay (Inpatient Stay Value Set).

See the corresponding Excel document for the value sets referenced above. Follow the steps below to identify the eligible population:

Step 1: Identify all patients in the specified age range who had an outpatient visit

(Outpatient Value Set), an observation visit (Observation Value Set) or an ED visit

(ED Value Set) during the Intake Period (January 1–December 24 of the measurement year). with a diagnosis of acute bronchitis (Acute Bronchitis Value Set).

Do not include ED visits that result in an inpatient admission.

Step 2: Determine all acute bronchitis Episode Dates. For each patient identified in step 1, determine all outpatient or ED claims/encounters with a diagnosis of acute bronchitis.

Step 3: Test for Negative Comorbid Condition History. Exclude Episode Dates when the

patient had a claim/encounter with any diagnosis for a comorbid condition during the 12 months prior to or on the Episode Date. A code from any of the following meets criteria for a comorbid condition:

- HIV Value Set.
- Malignant Neoplasms Value Set.
- Emphysema Value Set.
- COPD Value Set.
- Cystic Fibrosis Value Set.
- Comorbid Conditions Value Set.

Step 4: Test for Negative Medication History. Exclude Episode Dates where a new or refill prescription for an antibiotic medication (Table AAB-D) was filled 30 days prior to the Episode Date or was active on the Episode Date.

Step 5: Test for Negative Competing Diagnosis. Exclude Episode Dates where during the period 30 days prior to the Episode Date through 7 days after the Episode Date (inclusive) the patient had a claim/encounter with any competing diagnosis. A code from either of the following meets criteria for a competing diagnosis:

- Pharyngitis Value Set.
- Competing Diagnosis Value Set.

Step 6: Select the index episode start date. This measure examines the earliest eligible episode per patient.

(See the corresponding Excel document for the value sets referenced above)

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

As listed in the denominator details, the final denominator population does not include episodes with a history of select comorbid conditions, history of antibiotic use, or presence of a competing diagnosis

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

The measure excludes episodes with the following comorbid conditions during the 12 months prior to or on the Episode Date. A code from any of the following meets criteria for a comorbid condition:

- HIV Value Set.

- Malignant Neoplasms Value Set.
- Emphysema Value Set.
- COPD Value Set.
- Cystic Fibrosis Value Set.
- Comorbid Conditions Value Set.

The measure excludes episode with a new or refill prescription for an antibiotic medication (Table AAB-D) was filled 30 days prior to the Episode Date or was active on the Episode Date.

The measure excludes episodes with the following competing diagnoses during the period 30 days prior to the Episode Date through 7 days after the Episode Date (inclusive) the patient had a claim/encounter with any competing diagnosis. A code from either of the following meets criteria for a competing diagnosis:

- Pharyngitis Value Set.
- Competing Diagnosis Value Set.

See the corresponding Excel document for the value sets referenced above.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

HEDIS data are stratified by plan type (i.e. commercial, Medicaid). For this measure, a total rate is reported, along with three age stratifications (3 months–17 years; 18–64 years; 65 years and older).

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Other (specify):

If other: The measure is reported as an inverted rate [1 – (numerator/denominator)], therefore a higher score represents the proportion of episodes for which antibiotics were not prescribed.

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

Step 1: Identify all members who had an outpatient visit (Outpatient Value Set), a telephone visit (Telephone Visits Value Set), an e-visit or virtual check-in (Online Assessments Value Set), an observation visit (Observation Value Set) or an ED visit (ED Value Set) during the Intake Period, with a diagnosis of acute bronchitis/bronchiolitis (Acute Bronchitis Value Set).

Step 2: Determine all acute bronchitis/bronchiolitis Episode Dates. For each member identified in step 1, determine all outpatient, telephone, observation or ED visits, e-visits and virtual check-ins with a diagnosis of acute bronchitis/bronchiolitis.

Do not include visits that result in an inpatient stay (Inpatient Stay Value Set).

Step 3: Test for Negative Comorbid Condition History. Exclude Episode Dates when the member had a claim/encounter with any diagnosis for a comorbid condition during the 12 months prior to or on the Episode Date. A code from any of the following meets criteria for a comorbid condition:

- HIV Value Set.
- HIV Type 2 Value Set.
- Malignant Neoplasms Value Set.
- Other Malignant Neoplasm of Skin Value Set.
- Emphysema Value Set.
- COPD Value Set.
- Comorbid Conditions Value Set.
- Disorders of the Immune System Value Set.

Step 4: Test for Negative Medication History. Exclude Episode Dates where a new or refill prescription for an antibiotic medication (AAB Antibiotic Medications List) was filled 30 days prior to the Episode Date or was active on the Episode Date.

Step 5: Test for Negative Competing Diagnosis. Exclude Episode Dates where the member had a claim/encounter with a competing diagnosis on or 3 days after the Episode Date. A code from either of the following meets criteria for a competing diagnosis:

- Pharyngitis Value Set.
- Competing Diagnosis Value Set.

Step 6: Calculate continuous enrollment. The member must be continuously enrolled without a gap in coverage from 30 days prior to the Episode Date through 3 days after the Episode Date (34 total days).

Step 7: Deduplicate eligible episodes. If a member has more than one eligible episode in a 31-day period, include only the first eligible episode. For example, if a member has an eligible episode on January 1, include the January 1 visit and do not include eligible episodes that occur on or between January 2 and January 31; then, if applicable, include the next eligible episode that occurs on or after February 1. Identify visits chronologically, including only one per 31-day period.

Note: The denominator for this measure is based on episodes, not on members. All eligible episodes that were not excluded or deduplicated remain in the denominator.

Step 8: Calculate the numerator. Determine the number of events in the eligible population with a dispensed antibiotic medication on or three days after the episode date.

Step 9: Calculate a rate (number of antibiotics/eligible population).

Step 10: Subtract the rate calculated in step 9 from one to invert the measure result to represent appropriate treatment for acute bronchitis/bronchiolitis (i.e., antibiotic not prescribed). The measure is reported as an inverted rate (i.e., 1 – numerator/denominator) to reflect the number of episodes not associated with a dispensed antibiotic (higher is better).

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

N/A

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Claims

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from health plans via the Interactive Data Submission System (IDSS) portal.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Health Plan

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Emergency Department and Services, Outpatient Services

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

N/A

2. Validity – See attached Measure Testing Submission Form

AAB_0058_Testing_Form-637412883515387767.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

No - This measure is not risk-adjusted

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 0058
Measure Title: Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis
Date of Submission: 11/9/2020

Type of Measure:

Measure	Measure (continued)
☐ Outcome (*including PRO-PM*)	☐ Composite – *STOP – use composite testing form*
☐ Intermediate Clinical Outcome	☐ Cost/resource
☒ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	*

*cell intentionally left blank

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

2020 Submission

This measure was tested using administrative claims data from Medicaid and commercial plans nationwide that reported data for the annual Healthcare Effectiveness Data and Information Set (HEDIS)

2012 Submission

This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via the Interactive Data Submission System (IDSS) portal.

1.3. What are the dates of the data used in testing?

2020 Submission

January 1, 2019 through December 31, 2019

2012 Submission

2009, 2010, 2011

1.4. What levels of analysis were tested? (*testing must be provided for* **all** *the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other:	☐ other:

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

2020 Submission

This measure was recently expanded to include members 3 months of age and older (previously limited to adults 18-64 years of age). The measure assesses the percentage of episodes for members 3 months of age and older with a diagnosis of acute bronchitis/bronchiolitis that did not result in an antibiotic dispensing event. This measure is reported as an inverted rate [1-(numerator/eligible population)]. A higher rate indicates appropriate treatment for acute bronchitis/bronchiolitis (i.e., episodes that did *not* result in an antibiotic dispensing event).

The measure is reported at the health plan level and includes members enrolled in commercial, Medicaid and Medicare lines of business. Testing was therefore performed at the health plan level.

Measure score reliability and construct validity testing

Data used to assess reliability and validity were calculated from all reporting commercial health and Medicaid plans for this measure. There were 406 commercial plans and 213 Medicaid plans reporting this measure for the 2019 measurement year. The plans were geographically diverse and varied in size.

Systematic evaluation of face validity

The measure was assessed for face validity through two independent panels of experts:

- NCQA's Committee on Performance Measurement (CPM) oversees measures used in NCQA programs and includes representation by purchasers, consumers, health plans, health care providers, and policy makers. This panel is composed of 17 independent members that reflect the diversity of constituencies that performance measurement serves. The CPM's recommendations are reviewed and approved by NCQA's Board of Directors.
- NCQA's Antibiotic Overuse Measurement Advisory Panel is composed of 8 independent members representing hospitals, public policy research, public health and universities. This panel oversees HEDIS antibiotic use measures to align with current evidence-based guidelines and practices.

2012 Submission

This measure assesses the percentage of adults 18-64 years of age with a diagnosis of acute bronchitis who were not dispensed an antibiotic prescription.

Measure score reliability

The data exist in HEDIS Performance Measurement data for 2011 which include 411 commercial plans and 145 Medicaid plans. The health plans were a geographically diverse sample.

Systematic evaluation of face validity

The Avoidance of Antibiotic Treatment in Adults with Acute Bronchitis/Bronchiolitis measure was tested for face validity with panels of experts, both advisory panels and other subject matter workgroups to provide the clinical and technical knowledge required to develop the measure. The Adult Antibiotic Expert Panel included 20 experts with representation by consumers, health plans, health care providers and policy makers. Additional HEDIS Expert Panels and the Technical Measurement Advisory Panel (TMAP) provide invaluable assistance by identifying methodological issues and giving feedback. NCQA's Committee on Performance Measurement (CPM) is made up of 21 members reflecting the diversity of constituencies that performance measurement serves; members bring other perspectives and additional expertise in quality management and the science of measurement. The CPM meets with the NCQA Board of Directors to recommend measures for inclusion in HEDIS.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

2020 Submission

HEDIS data are summarized at the health plan level and stratified by plan type (i.e. commercial, Medicaid). The eligible population for this measure is based on episodes. NCQA does not collect patient-level data from health plans but can provide age/sex information.

Table 1. Median eligible population for *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* by plan type, calendar year 2019 data

Product Type	Number of Plans	Median number of eligible episodes per plan
Commercial	406	1,208
Medicaid	213	3,467

2012 Submission

Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

The analysis included a geographically diverse universe of commercial and Medicaid plans between 2009 and 2011.

Number of commercial health plans, 2009= 426

Number of commercial health plans, 2010= 422

Number of commercial health plans, 2011= 404

Number of Medicaid health plans, 2009= 112

Number of Medicaid health plans, 2010= 133

Number of Medicaid health plans, 2011= 145

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

2020 Submission

There were no differences in the data used for reliability, construct validity or meaningful differences in performance testing. As described above in Section 1.5, two multi-stakeholder expert panels assessed face validity.

2012 Submission

N/A

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

2020 Submission

We did not assess data by social risk factors. Social risk factor data were not available in reported results. This measure is specified for Medicaid, commercial and Medicare members aged 3 months and older. NCQA is actively engaged with partners including the CMS Office of Minority Health in identifying feasible methods to further integrate social risk factors into health plan quality measures, with a focus on stratification. This is aligned with recent recommendations from MedPAC and ASPE on optimal methods for addressing social risk in quality measurement and programs.[1,2]This is an NCQA wide initiative. Our intent is to implement methods to bridge data concerns in the future.

1. Medicare Payment Advisory Commission. (2020). The Medicare Advantage program: Status report. In Report to the Congress: Medicare Payment Policy (p. 397). http://medpac.gov/docs/default-source/reports/mar20_medpac_ch13_sec.pdf
2. Office of the Assistant Secretary for Planning and Evaluation, & U.S. Department of Health & Human Services. (2020). Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-based-purchasing-programs-reports

2012 Submission

The measure is not stratified to detect disparities. NCQA has participated with IOM and others in attempting to include information on disparities in measure data collection. However, at the present time, these data, at all levels (claims data, paper chart review, and electronic records), are not coded in a standard manner, and are incompletely captured. There are no consistent standards for what entity (physician, group, plan, employer) should capture and report these data. While requiring data reporting could push the field forward, doing so could create a substantial burden with inability to use the data because of its inconsistency. Currently, we agree with the IOM report that disparities are best considered by the use of zip code analysis which has limited applicability in most reporting situations. At the health plan level, for HEDIS health plan data collection, NCQA has extensive data related to the use of stratification by insurance status (Medicare, Medicaid and private-commercial) and would strongly recommend this process where the data base supporting the measurement includes this information. However, we believe that the measure specifications should NOT require this since the measure is still useful where the data needed to determine disparities cannot be ascertained from the data available.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

2020 Submission

We utilized the methodology described by John Adams (Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009) to calculate signal-to-noise reliability. This methodology uses the Beta-binomial model to assess how well one can confidently distinguish the performance of one reporting entity from another. Conceptually, the Beta-binomial model is the ratio of signal to noise. The signal is the proportion of the variability in measured performance that can be explained by real differences across reporting entities (plans, physicians, etc.) in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures, such as the *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* measure. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance across reporting entities.

For the *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* measure, commercial and Medicaid health plans are the reporting entities. For the formulas and explanations below, we use these health plans as the reporting entities.

The formula for signal-to-noise reliability is:

$$\text{Signal-to-noise reliability} = \sigma^2_{\text{plan-to-plan}} / (\sigma^2_{\text{plan-to-plan}} + \sigma^2_{\text{error}})$$

Therefore, we need to estimate two variances: 1) variance between plans ($\sigma^2_{\text{plan-to-plan}}$); 2) variance within plans (σ^2_{error}).

1. Variance between plans = $\sigma^2_{\text{plan-to-plan}} = (\alpha\ \beta) / (\alpha + \beta + 1)(\alpha + \beta)^2$

   α and β are two shape parameters of the Beta-Binomial distribution, $\alpha > 0$, $\beta > 0$

2. Variance within plans: $\sigma^2_{\text{error}} = \hat{p}(1- \hat{p})/n$

   $\hat{p}$ = observed rate for the plan

   n = plan-specific denominator for the observed rate (in this case, the number of eligible acute bronchitis/bronchiolitis episodes)

Using Adams' 2009 methodology, we estimated the reliability for each reporting entity, then averaged these reliability estimates across all reporting entities to produce a point estimate of signal-to-noise reliability. We label this point estimate "mean signal-to-noise reliability". The mean signal-to-noise reliability measures how well, on average, the measure can differentiate between reporting entity performance on the measure.

Along with the point estimate of mean signal-to-noise reliability, we are also providing:

1. The standard error (SE) and 95% confidence interval (95% CI) of the mean signal-to-noise reliability for all plans and stratified by the denominator size (number of eligible episodes per plan) in Table 3. The SE and 95% CI of the mean signal-to-noise reliability provides information about the stability of reliability. The 95% CI is the mean signal-to-noise reliability ± (1.96*SE). The narrower the confidence interval, the less the mean signal-to-noise reliability estimate will change due to idiosyncratic features of specific plans. We also stratified the results by the denominator size using terciles of the distribution to provide additional information about the stability of reliability.
2. The distribution (minimum, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, maximum) of the plan-level signal-to-noise reliability estimates. Each plan's reliability estimate is a ratio of signal to noise, as described above [$\sigma^2_{\text{plan-to-plan}} / (\sigma^2_{\text{plan-to-plan}} + \sigma^2_{\text{error}})$]. Variability between plans ($\sigma^2_{\text{plan-to-plan}}$) is the same for each plan, while the specific plan error (σ^2_{error}) varies. Reliability for each plan is an ordinal measure of how well one can determine where a given plan lies in the distribution of reliability across all plans, with higher estimates indicating better reliability. We also stratified the results by the denominator size using terciles of the distribution to provide additional information about the distribution of plan-level signal-to-noise reliability estimates. The number of plans in each stratum and the per-plan denominators of the performance rates are displayed in the summary tables.

This methodology allows us to estimate the reliability for each plan and summarize the distribution of these estimates.

2012 Submission

In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009) in work produced for the National Committee for Quality Assurance (NCQA). The following is quoted from the tutorial which focused on provider-level assessment: "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient." This approach is also relevant to health plans and other accountable entities.

Adams' approach uses a Beta-binomial model to estimate reliability; this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures. The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities.

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

2020 Submission

Table 2 shows the point estimates of mean signal-to-noise reliability using above methodology. The point estimate of mean signal-to-noise reliability at the health plan level ranges from 0.886 to 0.983.

Table 2. Point Estimates of Mean Signal-to-Noise Reliability by Product Type, Calendar Year 2019 Data

Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis	Point estimate: Mean Signal-To-Noise Reliability (Commercial)	Point estimate: Mean Signal-To-Noise Reliability (Medicaid)
Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (Total)	0.9634	0.9815
Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (age 3 Months-17 Years)	0.9216	0.9829
Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (age 18-64)	0.9539	0.9295
Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (age 65+)	0.8861	0.9609

Table 3 provides the point estimate of mean signal-to-noise reliability, its standard error, and the 95% CI for the *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (Total)* measure for commercial and Medicaid plans overall and stratified by the denominator size (distribution of the number of eligible episodes

per plan). Over all commercial plans, the reliability estimate is 0.963, and the 95% CI is (0.958, 0.969), indicating very good reliability. Stratified analyses show that reliability increase as plan size gets larger and stay above 0.9. Over all Medicaid plans, the reliability estimate is 0.982 and the 95% CI is (0.976, 0.987), indicating very good reliability. Results from the stratified analyses show that reliability exceeds 0.9 for all terciles.

Table 3. Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (Total)* Measure by Terciles of the Denominator Size and for All Submissions Stratified by Plan Type, Calendar Year 2019 Data

Stratification	Number of Plans	Number of Eligible Episodes per Plan (min - max)	Mean Signal-To-Noise Reliability	SE	95% CI
All Commercial	**406**	**30 - 37586**	**0.963**	**0.003**	**(0.958, 0.969)**
Tercile 1	135	30 – 636	0.927	0.0058	(0.915, 0.938)
Tercile 2	133	647 – 2435	0.984	0.0005	(0.983, 0.985)
Tercile 3	138	2455 - 37586	0.994	0.0003	(0.994, 0.995)
All Medicaid	**213**	**30 - 40245**	**0.982**	**0.003**	**(0.976, 0.987)**
Tercile 1	70	30 – 1680	0.970	0.0051	(0.959, 0.980)
Tercile 2	70	1854 – 6077	0.992	0.0003	(0.992, 0.993)
Tercile 3	73	6113 - 40245	0.995	0.0002	(0.995, 0.996)

SE: Standard Error of the mean.

95% CI: 95% confidence interval.

Table 4 summarizes the distribution of plan-level signal-to-noise reliability estimates for the *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (Total)* measure. Over all commercial plans, the estimates range from 0.585 to 0.99. The 10th percentile is 0.908 and the 50th percentile is 0.983, which exceed the 0.70 threshold for reliability. For Medicaid plans, the estimates range from 0.589 to 1.0; the 10th percentile is 0.96 and the 50th percentile is 0.994, indicating very good reliability. This table also include the distribution of plan-level signal-to-noise reliability estimates stratified by the tercile of the denominator size. Reliability estimates tend to be higher for plans with a larger denominator.

Table 4. Distribution of Plan-Level Signal-To-Noise Reliability for the Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (Total) Measure by Terciles of the Denominator Size and for All Submissions by Plan Type, Calendar Year 2019 Data

		Distribution of Plan Estimates of Signal-to-Noise Reliability						
Stratification	**Number of Plans**	**Min**	**P10**	**P25**	**P50**	**P75**	**P90**	**Max**
All Commercial	406	0.585	0.908	0.959	0.983	0.994	0.997	0.999
Tercile 1	135	0.636	0.847	0.911	0.957	0.967	0.972	0.974
Tercile 2	133	0.970	0.976	0.979	0.984	0.989	0.991	0.992
Tercile 3	138	0.987	0.990	0.991	0.994	0.997	0.998	0.999
All Medicaid	213	0.589	0.962	0.982	0.994	0.997	0.999	1.000
Tercile 1	70	0.702	0.935	0.974	0.984	0.988	0.990	0.991
Tercile 2	70	0.987	0.988	0.991	0.993	0.995	0.995	0.996

Stratification	Number of Plans	Min	P10	P25	P50	P75	P90	Max
Tercile 3	73	0.992	0.992	0.994	0.996	0.997	0.998	0.998

2012 Submission

Reliability statistics for this measure were calculated using HEDIS health plan performance data for 2011. The results are as follows:

Commercial

AAB - Reported rate 0.99

Medicaid

AAB -Reported rate 0.96

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

2020 Submission

Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. Both plan types had median reliability greater than 0.90 indicating that the measure has very good reliability.

Across all commercial plans, the reliability estimate is 0.963, and the 95% CI is (0.958, 0.969), indicating very good reliability. Stratified analyses show that reliability increase as plan size gets larger and stay above 0.9. Over all Medicaid plans, the reliability estimate is 0.982 and the 95% CI is (0.976, 0.987), indicating very good reliability. Results from the stratified analyses show that reliability exceeds 0.9 for all terciles.

2012 Submission

N/A

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

- ☒ **Empirical validity testing**
- ☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

2020 Submission

Construct Validity Testing of Performance Measure Score

NCQA performs Pearson correlation for construct validity using HEDIS health plan data. The test estimates the strength of linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a strong positive linear association: an increase in values of one variable is associated with increase in value of another variable. A value of 0 indicates no linear association. A value of -1

indicates a strong negative relationship in which an increase in values of the first variable is associated with a decrease in values of the second variable. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The sample size for the correlation analysis is the number of plans that reported both measures. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We adjusted our p-values to account for testing multiple correlations and used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

We tested for construct validity by exploring the following:

- Is *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* positively correlated with the HEDIS *Appropriate Treatment for Upper Respiratory Infection* measure which assesses the percentage of episodes of upper respiratory infection among members 3 months of age and older who were not dispensed an antibiotic prescription?
- Is *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* negatively correlated with the HEDIS *Antibiotic Utilization* measure which assesses the average number of outpatient antibiotic prescriptions per member per year?

We hypothesized that health plans with a high rate for the *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* measure would have a high rate for the *Appropriate Treatment for Upper Respiratory Infection* measure. Acute bronchitis/bronchiolitis and upper respiratory infections are largely driven by viruses and antibiotic stewardship efforts such as patient education and prior authorization target these common outpatient diagnoses to avoid inappropriate antibiotic prescribing.

We hypothesized that health plans with a high rate for the *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis* measure would have a low rate for the *Antibiotic Utilization* measure. At least 30% of outpatient antibiotic prescriptions are inappropriate and treatment for acute bronchitis/bronchiolitis is a large contributor to broad outpatient antibiotic overuse (CDC, 2019). Health plans with higher rates for appropriate antibiotic prescribing for acute bronchitis/bronchiolitis should tend to have lower overall antibiotic utilization.

Centers for Disease Control and Prevention. (2019). Antibiotic Use in the United States, 2018 Update: Progress and Opportunities. Atlanta, GA: US Department of Health and Human Services, CDC; 2019.

Systematic Assessment of Face Validity of Performance Measure Score

NCQA develops measures using a standardized process described below.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical measurement advisory panels (MAPs), whose members are authorities on clinical priorities for measurement, participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness, and feasibility. This information is gathered into a work-up format, which is vetted by the MAPs, the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. At this step, face validity is systematically determined by the CPM, which uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment. For the most recent updates to this measure in January 2019, all members of the CPM voted to approve moving forward with the proposed changes.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA about proposed new measures. Public comment offers an opportunity to assess the validity, feasibility, importance and other attributes of a measure from a wider audience. For this measure, a majority of public comment respondents supported the measure. NCQA MAPs and the technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. Face validity is then again

systematically assessed by the CPM. The CPM reviews all comments before making a final decision and votes to recommend approval of new measures for HEDIS. NCQA's Board of Directors then approves new measures.

2012 Submission

Construct Validity Testing of Performance Measure Score

N/A

Systematic Assessment of Face Validity of Performance Measure Score

NCQA identified and refined measure management into a standardized process called the HEDIS measure life cycle.

The following steps outline the components of the life cycle that are used to ensure that measure testing adheres to the highest standard possible.

*Step 1: Topic selection is the process of identifying measures that meet criteria consistent with the overall model for performance measurement. There is a huge universe of potential performance measures for future versions of HEDIS. The first step is identifying measures that meet formal criteria for further development.

NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs—whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness and feasibility. This information is gathered into a work-up format. Refer to What Makes a Measure "Desirable?" The work-up is vetted by NCQA's MAPs, and the TMAP, and various other panels.

*Step 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase.

Development includes the following tasks:

1. Ensure funding throughout measure testing
2. Prepare a detailed conceptual and operational work-up that includes a testing proposal
3. Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures.

The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

*Step 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to the CPM about new measures or about changes to existing measures.

NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM will be included in the next HEDIS year and reported as first-year measures.

*Step 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's Quality Compass or in accreditation scoring.

The first-year distinction guarantees that a measure can be efficiently collected, reported and audited before it is used for public accountability or accreditation. The purpose of this first-year distinction is to ensure that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues.

After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

*Step 5: Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be reported in Quality Compass and may be used for scoring in accreditation.

Step 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reevaluated at least every three years. NCQA staff continually

monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments contribute to measure evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, a third of the measurement set is researched for changes in clinical guidelines or health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume 2.

What makes a measure "Desirable"?

Whether considering the value of a new measure or the continuing worth of an existing one, we must define what makes a measure useful. HEDIS measures encourage improvement. The defining question for all performance measurement— "Where can measurement make a difference?"—can be answered only after considering many factors. NCQA has established three areas of desirable characteristics for HEDIS measures, discussed below.

1. Relevance: Measures should address features that apply to purchasers or consumers, or which will stimulate internal efforts toward quality improvement. More specifically, relevance includes the following attributes.

Meaningful: What is the significance of the measure to the different groups concerned with health care? Is the measure easily interpreted? Are the results meaningful to target audiences? Measures should be meaningful to at least one HEDIS audience (e.g., individual consumers, purchasers or health care systems). Decision makers should be able to understand a measure's clinical and economic significance.

Important to health: What is the prevalence and overall impact of the condition in the U.S. population? What significant health care aspects will the measure address? We should consider the type of measure (e.g., outcome or process), the prevalence of medical condition addressed by the measure and the seriousness of affected health outcomes.

Financially important: What financial implications result from actions evaluated by the measure? Does the measure relate to activities with high financial impact? Measures should relate to activities that have high financial impact.

Cost effective: What is the cost benefit of implementing the change in the health care system? Does the measure encourage the use of cost-effective activities or discourage the use of activities that have low cost-effectiveness? Measures should encourage the use of cost-effective activities or discourage the use of activities that have low cost-effectiveness.

Strategically important: What are the policy implications? Does the measure encourage activities that use resources efficiently? Measures should encourage activities that use resources most efficiently to maximize member health.

Controllable: What impact can the organization have on the condition or disease? What impact can the organization have on the measure? Health care systems should be able to improve their performance. For outcome measures, at least one process should be controlled and have an important effect on outcome. For process measures, there should be a strong link between the process and desired outcome.

Variation across systems: Will there be variation across systems? There should be the potential for wide variation across systems. Potential for improvement: Will organizations be able to improve performance? There should be substantial room for performance improvement.

2. Scientific soundness: Perhaps in no other industry is scientific soundness as important as in health care. Scientific soundness must be a core value of our health care system—a system that has extended and improved the lives of countless individuals.

Clinical evidence: Is there strong evidence to support the measure? Are there published guidelines for the condition? Do the guidelines discuss aspects of the measure? Does evidence document a link between clinical

processes and outcomes addressed by the measure? There should be evidence documenting a link between clinical processes and outcomes.

Reproducible: Are results consistent? Measures should produce the same results when repeated in the same population and setting.

Valid: Does the measure make sense? Measures should make sense logically and clinically, and should correlate well with other measures of the same aspects of care.

Accurate: How well does the measure evaluate what is happening? Measures should precisely evaluate what is actually happening.

Risk adjustment: Is it appropriate to stratify the measure by age or another variable? Measure variables should not differ appreciably beyond the health care system's control, or variables should be known and measurable. Risk stratification or a validated model for calculating an adjusted result can be used for measures with confounding variables.

Comparability of data sources: How do different systems affect accuracy, reproducibility and validity? Accuracy, reproducibility and validity should not be affected if different systems use different data sources for a measure.

3. Feasibility:

The goal is not only to include feasible measures, but also to catalyze a process whereby relevant measures can be made feasible.

Precise specifications: Are there clear specifications for data sources and methods for data collection and reporting? Measures should have clear specifications for data sources and methods for data collection and reporting.

Reasonable cost: Does the measure impose a burden on health care systems? Measures should not impose an inappropriate burden on health care systems.

Confidentiality: Does data collection meet accepted standards of member confidentiality? Data collection should not violate accepted standards of member confidentiality.

Logistical feasibility: Are the required data available?

Auditability: Is the measure susceptible to exploitation or "gaming" that would be undetectable in an audit? Measures should not be susceptible to manipulation that would be undetectable in an audit.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

2020 Submission

Construct validity testing

Table 5. Pearson Correlation Coefficient for Commercial and Medicaid health plans for the *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (Total)* Measure, Calendar Year 2019 Data

Measure	Appropriate Treatment for Upper Respiratory Infection (Total)	Antibiotic Utilization - Average Scrips for Antibiotics PMPY (M/F)
Commercial	0.68	-0.64
(N=, p value =)	(396, $p < 0.001$)	(386, $p < 0.001$)
Medicaid	0.68	-0.60
(N=, p value =)	(212, $p < 0.001$)	(192, $p < 0.001$)

N = number of health plans reporting both measures

Face validity testing

Input from our multi-stakeholder measurement advisory panels and those submitting to public comment indicate the measure has face validity.

2012 Submission

Step 1: The Avoidance of Antibiotic Treatment in Adults with Acute Bronchitis measure was developed to address a gap in care concerning the need to decrease excess antibiotic use in ambulatory practice, fueled by the epidemic increase in antibiotic resistant Streptococcus pneumonia. NCQA's Performance Measurement Department and the Respiratory MAP worked together to determine the most appropriate way to decrease antibiotic use.

Step 2: The measure was written, field-tested, and presented to the CPM in 2005. The CPM recommended sending the measure to public comment.

Step 3: The measure was released for Public Comment in spring 2005. We received and responded to comments on this measure. The CPM approved to move this measure to first year data collection. The voting process involved a simple majority vote with a quorum of CPM members.

Step 4: The Avoidance of Antibiotic Treatment in Adults with Acute Bronchitis measure was introduced in HEDIS 2007. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure public reporting. The voting process involved a simple majority vote with a quorum of CPM members.

Step 5: The Avoidance of Antibiotic Treatment in Adults with Acute Bronchitis measure will be reevaluated in 2013. POTENTIAL THREATS TO VALIDITY: This measure is validly specified by excluding patients from the denominator who may have conditions where antibiotics may be warranted, such as chronic comorbidities or competing bacterial infections. Field-test results show significantly lower rates of antibiotic prescribing found in patients with comorbidities (such as COPD) and with competing diagnoses where antibiotics may be indicated. Medical record validation of plans' administrative data was conducted to demonstrate the validity of administrative data to accurately identify the denominator population and exclusions, as well as the reliability of pharmacy data to capture inappropriate antibiotic prescriptions. Overall, concordance of administrative data with medical record document shows that the denominator specifications (using ICD-9 code 466.0) are highly reliable and accurate (about 90 percent) in identifying patients with acute bronchitis. Findings suggest that any unintentional inclusion of patients with comorbidities not identified in the administrative data (about 20 percent according to medical record documentation) or competing diagnoses (about 15 percent) the measure denominator would not adversely impact a plan's performance, and in fact, under-estimates the true extent of inappropriate antibiotic prescribing. Actual antibiotic prescribing rates (prescriptions ordered) may in fact be higher (by about 10 percent) than indicated by administrative data, since administrative data only captures filled prescriptions.

POTENTIAL THREATS TO VALIDITY: This measure is validly specified by excluding patients from the denominator who may have conditions where antibiotics may be warranted, such as chronic comorbidities or competing bacterial infections. Field-test results show significantly lower rates of antibiotic prescribing found in patients with comorbidities (such as COPD) and with competing diagnoses where antibiotics may be indicated. Medical record validation of plans' administrative data was conducted to demonstrate the validity of administrative data to accurately identify the denominator population and exclusions, as well as the reliability of pharmacy data to capture inappropriate antibiotic prescriptions. Overall, concordance of administrative data with medical record document shows that the denominator specifications (using ICD-9 code 466.0) are highly reliable and accurate (about 90 percent) in identifying patients with acute bronchitis. Findings suggest that any unintentional inclusion of patients with comorbidities not identified in the administrative data (about 20 percent according to medical record documentation) or competing diagnoses (about 15 percent) the measure denominator would not adversely impact a plan's performance, and in fact, under-estimates the true extent of inappropriate antibiotic prescribing. Actual antibiotic prescribing rates (prescriptions ordered) may in fact be higher (by about 10 percent) than indicated by administrative data, since administrative data only captures filled prescriptions.

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

2020 Submission

Commercial:

Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis is positively correlated with *Appropriate Treatment for Upper Respiratory Infection* (correlation coefficient = 0.68, p < 0.001) and is negatively correlated with *Antibiotic Utilization* (correlation = -0.64, p < 0.001).

Medicaid:

Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis is positively correlated with *Appropriate Treatment for Upper Respiratory Infection* (correlation coefficient = 0.679, p < 0.001) and is negatively correlated with *Antibiotic Utilization* (correlation = -0.598, p < 0.001).

2012 Submission

N/A

_____

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b3*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2020 Submission

NCQA does not collect exclusion data during annual HEDIS reporting. Individuals requiring hospice services during the measurement year are consistently excluded across HEDIS measures as advised by clinical Measurement Advisory Panels. The competing diagnosis, comorbid condition and negative antibiotic history exclusions for the measure are recommended by NCQA's Antibiotic Overuse Measurement Advisory Panel. These exclusions were assessed during development. Initial measure field testing was conducted across three national health plan organizations reflecting all lines of business (Medicare, Medicaid and commercial) using enrollment and claims data. The raw frequency of exclusion events and proportion of denominator events excluded were evaluated.

The steps to conduct this testing are described below:

1. NCQA recruited three health plan organizations with Medicare, Medicaid and commercial product lines to participate in field testing. These sites provided relevant data on their member population as well as qualitative information on their experience of collecting and reporting antibiotic use information.
2. The NCQA team developed a standardized data collection protocol based on a uniform data model developed for the specific purpose of collecting standardized, electronic clinical data. Each plan was asked to submit an aggregate table of overall plan descriptive information as well as a member-level comma-separated value (csv) file containing all the requested elements of the data model.
3. Using the csv file submitted by each plan, NCQA identified the eligible member population.
4. Among the eligible population, denominator events for each plan were identified following the logic:

   Step 1
   Identify all members who had an outpatient visit (Outpatient Value Set), a telephone visit (Telephone Visits Value Set), an online assessment e-visit or virtual check-in (Online Assessments Value Set) an observation visit (Observation Value Set) or an ED visit (ED Value Set) during the Intake Period, with a diagnosis of —customize to the measure form.

   Step 2
   Determine all URI Episode Dates. For each member identified in step 1, determine all outpatient, telephone, observation or ED visits, e-visits and virtual check-ins with a diagnosis of acute bronchitis/bronchiolitis (Acute Bronchitis Value Set).

   Exclude outpatient, ED or observation visits that result in an inpatient stay (Inpatient Stay Value Set).

5. Among the denominator events, NCQA evaluated the frequency of each exclusion and the proportion of total denominator events removed when independently implementing the exclusion.

The hospice exclusion is not tested individually, but rather implemented in all HEDIS measures based on expert panel feedback on clinical appropriateness.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2020 Submission

Table 6. Exclusion Analysis: Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis

Exclusion	Line of Business	Plan A: Total No. Visits	Plan A: No. Excluded	Plan A: Prop. Excluded	Plan B: Total No. Visits	Plan B: No. Excluded	Plan B: Prop. Excluded	Plan C: Total No. Visits	Plan C: No. Excluded	Plan C: Prop. Excluded
Competing Diagnosis (1)	Commercial	*	*	*	*	*	*	16,748	4,237	25.30%
Competing Diagnosis (2)	Medicaid	9052	1,389	15.30%	23,346	6,222	26.70%	11,344	2,503	22.10%
Competing Diagnosis (3)	Medicare	1270	115	9.10%	*	*	*	5,145	947	18.40%
Comorbid Condition (1)	Commercial	*	*	*	*	*	*	16,748	534	3.20%
Comorbid Condition (2)	Medicaid	9,052	859	9.50%	23,346	1,044	4.50%	11,344	359	3.20%
Comorbid Condition (3)	Medicare	1,270	291	22.90%	*	*	*	5,145	554	10.80%
Negative Medication History (1)	Commercial	*	*	*	*	*	*	16,748	2,592	15.50%
Negative Medication History (2)	Medicaid	9,052	1,435	15.90%	23,346	4,212	18.00%	11,344	2,377	21.00%

Exclusion	Line of Business	Plan A: Total No. Visits	Plan A: No. Excluded	Plan A: Prop. Excluded	Plan B: Total No. Visits	Plan B: No. Excluded	Plan B: Prop. Excluded	Plan C: Total No. Visits	Plan C: No. Excluded	Plan C: Prop. Excluded
Negative Medication History (3)	**Medicare**	1,270	243	19.10%	*	*	*	5,145	1,324	25.70%

*cell intentionally left blank

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis.* **Note***: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

2020 Submission

In order to capture unique outpatient visits for bronchitis/bronchiolitis that inappropriately result in an antibiotic prescribing event, it is necessary to exclude denominator events that are confounded by proximal outpatient events for competing diagnoses where an antibiotic is warranted (i.e. competing diagnosis such as a urinary tract infection). Additionally it is necessary to exclude denominator events that are confounded by previously ongoing antibiotic treatment within the prior 30 days as an antibiotic prescription during the window of evaluation may not be attributed to the denominator event for bronchitis/bronchiolitis, but instead a refill event for chronic antibiotic use or continued treatment for a previous infection (i.e. negative medication history). The competing diagnosis exclusion accounted for 25.3% of commercial events, ranged from 15.3% to 26.7% for Medicaid and ranged from 9.1% to 18.4% for Medicare. The negative antibiotic medication exclusion accounted for 15.5% of commercial events, ranged from 15.9% to 21.0% for Medicaid and ranged from 19.1% to 25.7% for Medicare. Clinical practice guidelines for acute bronchitis and bronchiolitis that recommend against antibiotic prescribing note different treatment considerations should be given to individuals with select immunocompromising comorbidities (i.e. comorbidity such as cystic fibrosis). These individuals should be excluded from the measure and testing indicated that this accounts for 3.2% of commercial events, ranged from 3.2% to 9.5% for Medicaid and ranged from 10.8% to 22.9% for Medicare.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of*

p<0.10; correlation of x or higher; patient factors should be present at the start of care) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

- ☐ **Published literature**
- ☐ **Internal data analysis**
- ☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b3.11. Optional Additional Testing for Risk Adjustment (**not required**, *but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

2020 Submission

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure.

To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the below 25th and above 75th percentile groups. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p-value of the test statistic is less than .05, then the two plans' performance is significantly different from each other.

2012 Submission

Comparison of means and percentiles.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g.,

number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined)

2020 Submission

Table 7. Variation in Performance for Commercial and Medicaid health plans for the *Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis (Total)* Measure, Calendar Year 2019 Data

Measure	N	Min	P10	P25	Mean	Median	P75	P90	Max	IQR	P value
Commercial	406	0.19	0.3	0.34	0.41	0.39	0.45	0.53	0.87	0.11	< 0.0001
Medicaid	213	0.29	0.4	0.45	0.52	0.51	0.58	0.65	1	0.13	< 0.0001

N = Number of plans reporting

IQR: Interquartile Range

p-value: the p-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile

2012 Submission

The following data are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. It includes number of health plans, percentiles, mean, min, max and standard deviations.

Data is summarized at the health plan level (i.e. "N" represents the number of health plans)

The rate is reported as an inverted rate (i.e. 1- numerator/denominator) to reflect the number of people in the health plans that were not dispensed an antibiotic.

Data is stratified by year and product line (i.e. commercial, Medicare, Medicaid)

Commercial

2011 RATE

N 404

Mean 22.03

StdDev 7.86

Min 8.47

P10 15.44

P25 17.74

P50 20.61

P75 23.94

Max 86.24

2010 RATE

N 422

Mean 23.37

StdDev 7.93

Min 12.77

P10 17.06

P25 19.04

P50 21.80

P75 25.26

Max 87.67

2009 RATE

N 426

Mean 25.48

StdDev 9.15
Min 9.89
P10 18.40
P25 20.31
P50 23.45
P75 27.39
Max 90.54
Medicaid
2011; RATE
N; 145
Mean; 23.57
StdDev; 7.71
Min; 11.91
P10; 15.09
P25; 18.78
P50; 22.15
P75; 26.23
Max; 54.76
2010 RATE
N 133
Mean 25.78
StdDev 10.41
Min 11.39
P10 16.79
P25 19.74
P50 23.56
P75 27.00
Max 66.98
2009; RATE
N; 112
Mean; 25.76
StdDev; 10.52
Min; 8.89
P10; 17.71
P25; 20.21
P50; 23.67
P75; 28.09
Max; 85.37

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities?
(i*.e., what do the results mean in terms of statistical and meaningful differences?*)

2020 Submission

There is a 0.11 gap in performance between Commercial plans at the 25th and 75th percentiles, a 0.13 gap in performance among Medicaid plans. The difference in performance between plans in the 25th percentile and 75th percentile is statistically significant.

2012 Submission

N/A

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

***If only one set of specifications, this section can be skipped*.**

This measure has only one set of specifications.

Note*: This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

2020 Submission

HEDIS measures apply to enrolled members in each health plan, and NCQA has a rigorous audit process to ensure the eligible population and numerator events for each measure are correctly identified and reported by the health plans. The audit process is designed to verify primary data sources used to populate measures and ensure specifications are correctly implemented.
The HEDIS Compliance Audit addresses the following functions:

- Information practices and control procedures
- Sampling methods and procedures
- Data integrity
- Compliance with HEDIS specifications
- Analytic file production
- Reporting and documentation

2012 Submission

N/A

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various*

rules for missing data/nonresponse; **if no empirical sensitivity analysis***, identify the approaches for handling missing data that were considered and pros and cons of each*)

2020 Submission

Missing data is addressed during the HEDIS audit process through NCQA-certified auditors assessing whether data sources are missing data by using standard audit methodologies. If it is found that a data source is missing data and the issue cannot be remedied, then the measure will be designated "materially biased" and the rate will not be used. After measures are added to the HEDIS volume, NCQA conducts a first-year analysis on the feasibility of the measure when widely implemented in the field. This includes assessing the rate of materially biased reporting in addition to other reporting issues such as small dominators. These issues are considered when approving a measure for public reporting.

2012 Submission

N/A

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted;* **if no empirical analysis***, provide rationale for the selected approach for missing data*)

2020 Submission

All commercial and Medicaid plans reporting for the HEDIS 2020 measurement year (calendar year 2019) were audited as described above. This means that the auditors did not find missing data sources for this measure and the rates are not materially biased.

2012 Submission

N/A

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic claims

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than

electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO's ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:

1) Information practices and control procedures
2) Sampling methods and procedures
3) Data integrity
4) Compliance with HEDIS specifications
5) Analytic file production
6) Reporting and documentation

In addition to the HEDIS Audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measures. This system is vital to the regular re-evaluation of the NCQA measures.

Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
*	Public Reporting NCQA Health Plan Ratings https://www.ncqa.org/hedis/reports-and-research/ratings-2020/ NCQA Annual State of Health Care Quality https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report Public Health/Disease Surveillance Centers for Disease Control and Prevention (CDC) Measuring Outpatient Antibiotic Prescribing https://www.cdc.gov/antibiotic-use/community/programs-measurement/measuring-antibiotic-prescribing.html Regulatory and Accreditation Programs NCQA Health Plan Accreditation https://www.ncqa.org/programs/health-plans/health-plan-accreditation-hpa/ Quality Improvement (external benchmarking to organizations) Align. Measure. Perform. Program (IHA) https://www.iha.org/sites/default/files/resources/my_2019_align._measure._perform._amp_manual_2019.pdf NCQA Annual State of Health Care Quality https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report NCQA Quality Compass https://www.ncqa.org/programs/data-and-information-technology/data-purchase-and-licensing/quality-compass/ Quality Improvement (Internal to the specific organization) CDC Core Elements of Outpatient Antibiotic Stewardship https://www.cdc.gov/antibiotic-use/community/pdfs/16_268900-A_CoreElementsOutpatient_508.pdf

*cell intentionally left blank

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

NCQA QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting health plans, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.
NCQA HEALTH PLAN RATING/REPORT CARDS: This measure is used to calculate health plan ratings which are reported on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. Due to COVID-19, NCQA will not release 2010-2021 Health plan ratings for any product line. However, in 2019, a total of 255 Medicare health plans, 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.
NCQA HEALTH PLAN ACCREDITATION: This program is a widely recognized, evidence-based program dedicated to quality improvement and measurement. It provides a comprehensive framework for organizations to align and improve operations in areas that are most important to states, employers and consumers. It's the only evaluation program that bases results on actual measurement of clinical performance (HEDIS® measures) and consumer experience (CAHPS® measures). As of October 2020, there are 507 commercial, 228 Medicare and 178 Medicaid health plans with accreditation, representing entities from all states and geographic regions.
NCQA STATE OF HEALTH CARE QUALITY: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care.?
INTEGRATED HEALTHCARE ASSOCIATION: The purpose is to provide comprehensive benchmarks and a reliable assessment of performance for medical groups, independent practice association (IPAs), and accountable care organizations (ACOs) across health plans. AMP Commercial HMO program now includes participation from eleven health plans and about 200 California physician organizations caring for over 9.5 million Californians enrolled in commercial HMO and point of service products—representing 90% of commercial HMO enrollment in the state. AMP's Medi-Cal—California's Medicaid program—now covers more than 13 million people, or approximately one in three Californians.
CDC MEASURING OUTPATIENT ANTIBIOTIC PRESCRIBING: Monitoring of outpatient antibiotic prescribing data is regularly conducted to analyze national and state antibiotic prescribing data in order to better understand trends in outpatient antibiotic prescribing, to identify where interventions to improve prescribing are most needed, and to measure progress. The CDC website lists average national performance on the HEDIS AAB measure. The CDC website are publicly available to all audiences.
CDC CORE ELEMENTS OF OUTPATIENT ANTIBIOTIC STEWARDSHIP: This program provides a framework for antibiotic stewardship for outpatient clinicians and facilities that routinely provide antibiotic treatment. One of the four core elements involves tracking and reporting of antibiotic prescribing and highlights the use of HEDIS AAB measure assessing overprescribing for bronchitis as a way organizations can monitor prescribing practices.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

Health plans that report HEDIS calculate their rates and know their performance when submitting to NCQA. NCQA publicly reports rates across all plans and also creates benchmarks in order to help plans understand how they perform relative to other plans. Public reporting and benchmarking are effective quality improvement methods.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

NCQA publishes HEDIS results annually in our Quality Compass tool. NCQA also presents data at various conferences and webinars. For example, at the annual HEDIS Update and Best Practices Conference (now the Quality Innovation Series), NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly and insight into new measure development projects. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c.1.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

NCQA measures are evaluated regularly using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System. This information enables NCQA to comprehensively assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.

NCQA released proposed measurement changes in our annual HEDIS Public Comment period in 2019, which is available to all audiences to provide feedback on proposed measure updates and changes. Advisory panels of experts in antibiotic overuse and infectious diseases were also consulted.

Proposed changes included the addition of the Medicare product line, the expansion of the age group to include members 3 months of age and older, and the transition to an episode-based denominator.

4a2.2.2. Summarize the feedback obtained from those being measured.

In general, health plans have not reported significant barriers to implementing this measure, as it uses the administrative data collection method. Questions have generally centered around minor clarification of the specifications, such as confirmation that information in claims meets the measure intent and questions about the supporting guidelines for the measure. Overall, NCQA heard support from the public for proposed measure updates.

4a2.2.3. Summarize the feedback obtained from other users

This measure has been deemed a priority measure by NCQA and other entities, as illustrated by its use in programs such as the Annual State of Healthcare Quality and the Health Plan Rating.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

We have provided minor clarifications about the measure during the annual update process in order to address questions received through the Policy Clarification Support System. NCQA considers feedback from the public, experts and other stakeholders when making decisions about updating measure specifications. As a result of the feedback we received, the proposed measure changes were implemented.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Given changes to the measure denominator and age ranges covered between MY 2018 and MY 2019, trends in performance cannot be assessed. For MY 2019, the average percentage of episodes that were not associated with a dispensed antibiotic was 40.8% for commercial plans and 33.7% for Medicaid plans (full performance distribution details in section 1b). These proportions indicate poor health plan performance on antibiotic prescribing for bronchitis and bronchiolitis, and substantiates continued use of the measure. With a national focus on antibiotic stewardship, the goal is for health plans to continue driving progress in appropriate and conservative antibiotic use.

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

There were no identified unexpected findings during testing or since implementation of this measure.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

0069 : Appropriate Treatment for Upper Respiratory Infection

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

Yes

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

N/A

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

No appendix Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): National Committee for Quality Assurance

Co.2 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Co.3 Measure Developer if different from Measure Steward: National Committee for Quality Assurance

Co.4 Point of Contact: Brittany, Wade, wade@ncqa.org, 202-530-0463-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

The Antibiotic Overuse Measurement Advisory Panel (AOMAP) advised NCQA during measure re-evaluation. They evaluated the way staff specified measures, assessed the content validity of measures and reviewed field test results. In addition to the AOMAP, NCQA also vetted these measures with a host of other stakeholders, which is a routine part of our measure development process. Thus, our measures are the result of consensus from a broad a diverse group of stakeholders, including the AOMAP technical experts. Our Committee on Performance Measurement (CPM) is a voting body that reviews evidence and input from stakeholders, and measures are only included in HEDIS if the CPM votes to do so.

Antibiotic Overuse Measurement Advisory Panel (AOMAP)

Diana Buist, MPH, PhD, Kaiser Permanente Washington Health Research Institute

Jonathan Finkelstein, MD, MPH, Boston Children's Hospital

Jeffrey Gerber, PhD, The Children's Hospital of Philadelphia

Catherine Gillespie, PhD, MPH, AARP Public Policy Institute

Jeffrey Linder, MD, MPH, Northwestern University

Karl Madaras-Kelly, PharmD, PMH, Idaho State University

Rita Mangione-Smith, MD, MPH, University of Washington

Dat Tran, MD, Oregon Public Health Division

Committee on Performance Measurement (CPM)

Andy Baskin, MD, CVS Health/Aetna

Elizabeth Drye, MD, SM, Yale School of Medicine

Mark Friedberg, MD,MPP, Blue Cross Blue Shield of Massachusetts

Andrea Gelzer, MD, MS, FACP, AmeriHealth Caritas

David Grossman, MD, MPH, Washington Permanente Medical Group

Christine S. Hunter, MD, RADM, MC, USN, Self-employed, Independent Board Director

David Kelley, MD, MPA, Chief Medical Officer, Pennsylvania Department of Human Services

Jeff Kelman, MD, MMSc., Chief Medical Officer, Center for Medicare Department of Health and Human Services (DHHS)

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, Self Employed

Amanda Parsons, MD, MBA, MetroPlus

Wayne Rawlins, MD, MBA, ConnectiCare

Misty Roberts, MSN, RN, CPHQ, PMP, Humana

Rudy Saenz, MD, MMM, GACOG, Riverside Medical Clinic

Marcus Thygeson, MD, MPH, Bind Benefits

JoAnn Volk, MA, Georgetown University, Center on Health Insurance Reforms

Rose Baez, RN, MSN, MBA, CPHQ, Blue Cross Blue Shield Association

Jeff Brady, MD, MPH, AHRQ

Ron Kline, MD, Office of Personnel Management

Danielle Lloyd, MPH, America's Health Insurance Plan (AHIP)

Chelsey Richards, MD, MPH, FACP, Centers for Disease Control and Prevention

Anecia Suneja, CNS-BC, Veterans Health Administration (VHA)

Sheri Winsper, RN, MSN, MSHA, National Quality Forum (NQF)

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2004

Ad.3 Month and Year of most recent revision: 05, 2019

Ad.4 What is your frequency for review/update of this measure? Approximately every 3 years, sooner if the clinical guidelines have changed significantly

Ad.5 When is the next scheduled review/update for this measure? 12, 2021

Ad.6 Copyright statement: © 2020 by the National Committee for Quality Assurance

1100 13th Street, NW, 3rd floor

Washington, DC 20005

Ad.7 Disclaimers: These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA. © 2012 by the National Committee for Quality Assurance